LMMSE Estimation and Interpolation of Continuous-Time Signals from Discrete-Time Samples Using Factor Graphs

Lukas Bolliger, Hans-Andrea Loeliger, and Christian Vogel

Abstract—The factor graph approach to discrete-time linear Gaussian state space models is well developed. The paper extends this approach to continuous-time linear systems / filters that are driven by white Gaussian noise. By Gaussian message passing, we then obtain MAP / MMSE / LMMSE estimates of the input signal, or of the state, or of the output signal from noisy observations of the output signal. These estimates may be obtained with arbitrary temporal resolution. The proposed input signal estimation does not seem to have appeared in the prior Kalman filtering literature.
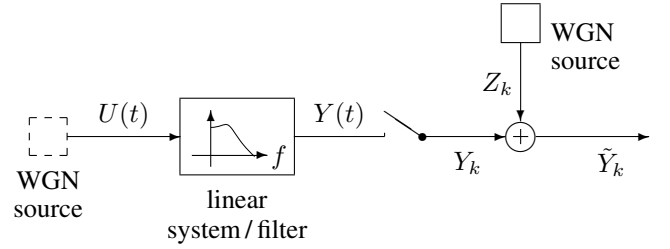

Fig. 1.   System model.

I. INTRODUCTION

Consider the system model shown in Fig. 1: a continuous-time linear time-invariant system / filter is fed by a continuous-time signal $U(t)$. The system output $Y(t)$ is sampled (at regular or irregular intervals) and the samples are corrupted by discrete-time additive white Gaussian noise. From the noisy samples $\tilde{Y}_k$, we wish to estimate the clean samples Y_k, or the clean signal $Y(t)$ at arbitrary instants t, or the state trajectory of the system, or—of particular interest in this paper—the input signal $U(t)$ at arbitrary instants t. We will not assume that any of these signals is bandlimited (in the strict sense required by the sampling theorem); instead, the key assumption in this paper is that the given linear system has a finite-dimensional state space representation.

Problems of this kind are ubiquitous. For example, Fig. 1 might model an analog-to-digital converter with a non-ideal anti-aliasing filter and with quantization noise Z_k; indeed, this application is a main motivation for this paper. As another example, Fig. 1 might model a sensor with some internal dynamics which limits its temporal resolution of the desired quantity $U(t)$. In both examples, we are primarily interested in estimating the input signal $U(t)$.

We will address these estimation problems under the further assumption that the input signal $U(t)$ is white Gaussian noise. It might perhaps seem at first that this assumption is problematic when $U(t)$ is actually the signal of interest, as in the two mentioned examples. However, we will argue that this assumption is meaningful in such cases and that the LMMSE

(linear minimum mean squared error) estimate of $U(t)$ is well defined and useful. An example of such an LMMSE estimate of $U(t)$ is shown in Fig. 11. The nature of this estimate will further be illuminated by reformulating it as a regularized least-squares problem with a penalty term $\int u(t)^2\,dt$, as will be discussed in Section V.

The assumption that $U(t)$ is white Gaussian noise turns our system model (Fig. 1) into a linear Gaussian model, and LMMSE estimation of the state trajectory or of the clean output signal $Y(t)$ amounts essentially to Kalman filtering[1] (or rather Kalman smoothing) [2]–[8]. However, estimation of the continuous-time input signal $U(t)$ does not seem to have been addressed in the Kalman filtering literature.

We will also consider some extensions of the system model including time-varying systems, vector signals, and systems with internal noise sources. These extensions are required in some of the motivating applications, but the extensions are straightforward and standard in Kalman filtering.

We will address these estimation problems (as described above) using factor graphs. Factor graphs [9]–[12] and similar graphical models [13]–[16] allow a unified description of system models and algorithms in many different fields. In particular, Gaussian message passing in factor graphs subsumes discrete-time Kalman filtering and many variations of it [9]–[12]. The graphical-model approach has facilitated the use of these techniques as components in more general inference problems and it has become a mode of teaching discrete-time Kalman filtering itself.

In this paper, we extend the factor graph approach to continuous-time models with discrete-time observations as

Lukas Bolliger and Hans-Andrea Loeliger are with the Dept. of Information Technology and Electrical Engineering, ETH Zurich, CH-8092 Zurich, Switzerland. Email: `loeliger@isi.ee.ethz.ch`, `lukas@bolligernet.ch`.

Christian Vogel is with the Telecomm. Research Center Vienna (FTW), Donau-City-Strasse 1, A-1220 Vienna, Austria. Email: `c.vogel@ieee.org`.

An abbreviated version of this paper was presented at the 2010 Information Theory & Appl. Workshop (ITA), La Jolla, CA, Feb. 2010 [1].

[1] Note that the Kalman-Bucy filter [3] addresses the different situation where the observations are continuous-time signals as well.

described above. This extension appears to be new[2], and it significantly enlarges the domain of graphical models. We note, in particular, that the LMMSE estimates of the continuous-time signals associated with such models (such as $U(t)$ and $Y(t)$ in Fig. 1) become computational objects that can be handled with arbitrary temporal resolution by Gaussian message passing.

Applications of the methods of this paper (in addition to those already mentioned) have been reported in [17] and [18]. In [17], a new method for sampling jitter correction is proposed that uses the slope of $Y(t)$, which is available in the state space model, in an iterative algorithm. In [18], a new approach to analog-to-digital conversion is proposed which combines unstable analog filters with digital estimation of $U(t)$ as proposed in the present paper. Both of these applications build on [1] (which does not contain the proofs) and rely on the present paper for a full justification of the proposed algorithms. Further applications (including beamforming with sensor arrays and Hilbert transforms) will be reported elsewhere.

In summary, this paper

- extends the factor graph approach to continuous-time models as in Fig. 1;
- extends Kalman smoothing (forward-backward Gaussian message passing) to the estimation of input signals as $U(t)$ in Fig. 1;
- provides the necessary background for subsequent work such as [17] and [18].

The paper builds on, and assumes some familiarity with, the factor graph approach to discrete-time Kalman filtering as given in [11].

The paper is structured as follows. The system model is formally stated in Section II and represented in factor graph notation in Section III. State estimation and output signal estimation are then essentially obvious, but some pertinent comments are given in Section IV. Estimation of the input signal is discussed in Section V. In Section VI, the estimation algorithms are illustrated by some simple numerical examples. A number of extensions of the basic system model are outlined in Section VII, and Section VIII concludes the paper.

The following notation will be used: $\overline{x}$ denotes the complex conjugate of x; A^T denotes the transpose of the matrix A; $A^\mathsf{H} \triangleq \overline{B}^\mathsf{T}$ denotes the Hermitian transpose of A; I denotes an identity matrix; "$\propto$" denotes equality up to a constant scale factor; $\mathcal{N}(m, \sigma^2)$ or $\mathcal{N}(m, V)$ denotes a normal (Gaussian) distribution with mean m and variance σ^2, or with mean vector m and covariance matrix V, respectively.

II. SYSTEM MODEL

Let $X \in \mathbb{R}^n$ be the state of a linear system (as, e.g., in Fig. 1) which evolves in time according to

$$\dot{X}(t) = AX(t) + bU(t) \tag{1}$$

where $\dot{X}$ denotes the derivative with respect to time and where both the matrix $A \in \mathbb{R}^{n \times n}$ and the vector $b \in \mathbb{R}^n$ are known.

[2]Another extension of graphical models to continuous time are continuous-time Bayesian networks [19]–[21], where the emphasis is on finite-state models and approximate inference. Yet another such extension is [22], where linear RLC circuits are described in terms of factor graphs.

The system output is the discrete-time signal $Y_1, Y_2, \ldots \in \mathbb{R}^\nu$ with

$$Y_k = CX(t_k) \tag{2}$$

where $t_1, t_2, \ldots \in \mathbb{R}$ (with $t_{k-1} < t_k$) are discrete instants of time and where $C \in \mathbb{R}^{\nu \times n}$ is known. We will usually observe only the noisy output signal $\tilde{Y}_1, \tilde{Y}_2, \ldots$ defined by

$$\tilde{Y}_k = Y_k + Z_k, \tag{3}$$

where $Z_1, Z_2, \ldots$ (the noise) are independent Gaussian random variables, each of which takes values in $\mathbb{R}^\nu$ and has a diagonal covariance matrix V_Z.

The (real and scalar) input signal $U(t)$ will be modeled as white Gaussian noise, i.e., for $t < t'$, the integral

$$\int_t^{t'} U(\tau) \, d\tau \tag{4}$$

is a zero-mean Gaussian random variable with variance $\sigma_U^2(t' - t)$, and any number of such integrals are independent random variables provided that the corresponding integration intervals are disjoint. In consequence, it is appropriate to replace (1) by

$$dX(t) = AX(t) \, dt + b \, U(t) \, dt \tag{5}$$

where $U(t) \, dt$ is a zero-mean Gaussian random variable with infinitesimal variance $\sigma_U^2 \, dt$.

As stated in the introduction, we will argue later (in Section V) that modeling $U(t)$ as white Gaussian noise is meaningful even when $U(t)$ is a (presumably smooth) signal of interest that we wish to estimate.

For any fixed initial state $X(t_0) = x(t_0)$, equation (5) induces a probability density $f(x(t_1)|x(t_0))$ over the possible values of $X(t_1)$ (where t_0 and t_1 are unrelated to the discrete times $\{t_k\}$ in (2)). Specifically, integrating (5) from $t = t_0$ to $t_1 > t_0$ yields

$$X(t_1) = e^{AT}X(t_0) + \int_0^T e^{A(T-\tau)}bU(t_0 + \tau) \, d\tau \tag{6}$$

with $T \triangleq t_1 - t_0 > 0$. If $U(t)$ is white Gaussian noise (with σ_U^2 as above), then the integral in (6) is a zero-mean Gaussian random vector with covariance matrix[3] [23]–[25]

$$V_S = \sigma_U^2 \int_0^T e^{A(T-\tau)}bb^\mathsf{T}(e^{A(T-\tau)})^\mathsf{T} \, d\tau \tag{7}$$

$$= \sigma_U^2 \int_0^T e^{A\tau}bb^\mathsf{T}(e^{A\tau})^\mathsf{T} \, d\tau. \tag{8}$$

It is thus clear that, for fixed $X(t_0) = x(t_0)$, $X(t_1)$ is a Gaussian random vector with mean $e^{AT}x(t_0)$ and covariance matrix V_S, i.e.,

$$f(x(t_1)|x(t_0)) \propto e^{-\frac{1}{2}(x(t_1) - e^{AT}x(t_0))^\mathsf{T} V_S^{-1}(x(t_1) - e^{AT}x(t_0))}. \tag{9}$$

[3]This covariance matrix is essentially the controllability Gramian. However, controllability is not required in this paper.
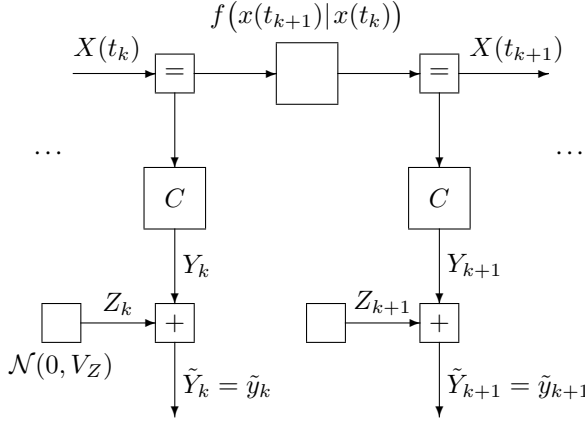
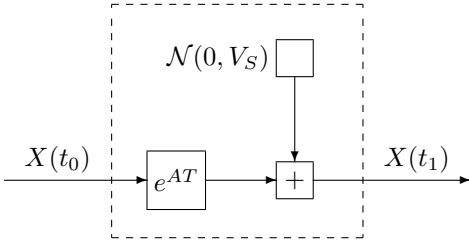Fig. 2. Factor graph of the system model with observations $\tilde{Y}_k = \tilde{y}_k$.

Fig. 3. Factor graph of $f\big(x(t_1)|x(t_0)\big)$ according to (6)–(8). (The values of t_0 and t_1 are not restricted to the discrete times $\{t_k\}$ in Fig. 2.)

III. FACTOR GRAPH OF SYSTEM MODEL

We will use Forney factor graphs (also known as normal factor graphs [26]) as in [10] and [11]. The nodes / boxes in such a factor graph represent factors and the edges in the graph represent variables.

In this notation, the system model of Section II may be represented by the factor graph shown in Fig. 2. More precisely, Fig. 2 represents the joint probability density of the variables in the system model at discrete times $t_1, t_2, \ldots$.

Note that Fig. 2 shows only a section (from t_k to t_{k+1}) of the factor graph; the complete factor graph starts at time t_0 and ends at some time t_K, and it may contain additional nodes to represent any pertinent initial or final conditions. Note also that, apart from the Gaussian nodes / factors $f\big(x(t_{k+1})|x(t_k)\big)$ and $\mathcal{N}(0, V_Z)$, the nodes / boxes in Fig. 2 represent linear constraints.

For details of this factor graph notation, we refer to [11].

As shown in Fig. 2, the function (9) can immediately be used as a node in a factor graph. However, the function (9) can itself be represented by nontrivial factor graphs. A first such factor graph is shown in Fig. 3, which corresponds to (6)–(8). Plugging Fig. 3 into Fig. 2 results in a standard discrete-time linear Gaussian factor graph as discussed in depth in [11].

The factor graph of Fig. 2 is easily refined to arbitrary temporal resolution by splitting the node / factor $f\big(x(t_{k+1})|x(t_k)\big)$ as shown in Fig. 4. In this way, both the state $X(t)$ and the output signal $Y(t) = CX(t)$ become available for arbitrary instants t between t_k and t_{k+1}.
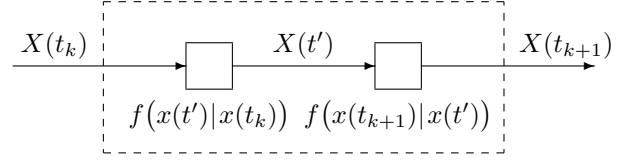

Fig. 4. Splitting the node / factor $f\big(x(t_{k+1})|x(t_k)\big)$ to access the state at an intermediate point in time t'.

Each of the factors in Fig. 4 can, of course, be replaced by the corresponding decomposition according to Fig. 3.

Note that the input signal $U(t)$ is not explicitly represented in Figures 2–4. For the estimation of $U(t)$, we will therefore need another decomposition of the node / factor $f\big(x(t_{k+1})|x(t_k)\big)$.

IV. GAUSSIAN MESSAGE PASSING, STATE ESTIMATION, AND OUTPUT SIGNAL ESTIMATION

Having thus obtained a discrete-time factor graph (with an arbitrary temporal resolution), estimating $X(t)$ or $Y(t)$ from the noisy observations $\tilde{Y}_1 = \tilde{y}_1$, $\tilde{Y}_2 = \tilde{y}_2$, … by means of Gaussian message passing is standard and discussed in detail in [11] (cf. also [10] and [12]). We therefore confine ourselves to a few general remarks (mostly excerpted from [10] and [11]) and some additional remarks on message passing through the node / factor $f\big(x(t)|x(t_0)\big)$.
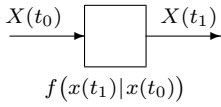
A. General Remarks

1) In linear Gaussian factor graphs such as Figures 2–4 (where all factors are either Gaussians or linear constraints), all sum-product messages are Gaussians and sum-product message passing coincides with max-product message passing. Moreover, MAP (maximum *a posteriori*) estimation coincides both with MMSE (minimum mean squared error) estimation and with LMMSE (linear / affine MMSE) estimation.

2) In general, every edge in the factor graph carries two messages, one in each direction. Since all the edges in Figures 2–4 are directed (i.e., drawn with an arrow), we can unambiguously refer to the forward message $\overrightarrow{\mu}_X$ and the backward message $\overleftarrow{\mu}_X$ along the edge representing some variable X.

3) Gaussian messages have the form

$$\mu(x) \propto e^{-\frac{1}{2}(x-m)^{\mathsf{T}}W(x-m)}; \qquad (10)$$

they are naturally parameterized by the mean vector m and either the matrix W or the covariance matrix $V\ (= W^{-1})$. Degenerate Gaussians, where either W or V do not have full rank, are often permitted and sometimes unavoidable; in such cases, only W or V, but not both, are well defined. We will use the symbols $\overrightarrow{m}_X$ and $\overrightarrow{V}_X$ (or $\overrightarrow{W}_X$) to denote the parameters of the forward message (along some edge / variable X) and $\overleftarrow{m}_X$ and $\overleftarrow{V}_X$ (or $\overleftarrow{W}_X$) for the parameters of the backward message.

4) The natural scheduling of the message computations in Fig. 2 consists of a forward recursion for $\overrightarrow{\mu}_{X(t_k)}$ and

$$\overrightarrow{m}_{X(t_1)} = e^{A(t_1-t_0)}\overrightarrow{m}_{X(t_0)} \qquad (\text{I.1})$$

$$\overrightarrow{V}_{X(t_1)} = e^{A(t_1-t_0)}\overrightarrow{V}_{X(t_0)}e^{A^\mathsf{T}(t_1-t_0)}$$
$$+ \sigma_U^2 \underbrace{\int_0^{t_1-t_0} e^{A\tau}bb^\mathsf{T}e^{A^\mathsf{T}\tau}d\tau}_{Q\overrightarrow{\Theta}(t_1-t_0)Q^\mathsf{H}\ \text{see (12)}} \qquad (\text{I.2})$$

$$\overleftarrow{m}_{X(t_0)} = e^{-A(t_1-t_0)}\overleftarrow{m}_{X(t_1)} \qquad (\text{I.3})$$

$$\overleftarrow{V}_{X(t_0)} = e^{-A(t_1-t_0)}\overleftarrow{V}_{X(t_1)}e^{-A^\mathsf{T}(t_1-t_0)}$$
$$+ \sigma_U^2 \underbrace{\int_0^{t_1-t_0} e^{-A\tau}bb^\mathsf{T}e^{-A^\mathsf{T}\tau}d\tau}_{Q\overleftarrow{\Theta}(t_1-t_0)Q^\mathsf{H}\ \text{see (14)}} \qquad (\text{I.4})$$

$$\hat{u}(t) = \sigma_U^2 b^\mathsf{T}\left(\overleftarrow{V}_{X(t)} + \overrightarrow{V}_{X(t)}\right)^{-1}\left(\overleftarrow{m}_{X(t)} - \overrightarrow{m}_{X(t)}\right) \qquad (\text{I.5})$$

an independent backward recursion[4] for $\overleftarrow{\mu}_{X(t_k)}$. Both of these recursions use the messages $\overleftarrow{\mu}_{Y_k}$ with parameters $\overleftarrow{m}_{Y_k} = \tilde{y}_k$ and $\overleftarrow{W}_{Y_k} = V_z^{-1}$ (assuming $\tilde{Y}_k = \tilde{y}_k$ is known; if $\tilde{Y}_k$ is not observed / unknown, then $\overleftarrow{W}_{Y_k} = 0$ and $\overleftarrow{\mu}_{Y_k}(y_k) = 1$).

5) Since the factor graph in Fig. 2 has no cycles, the *a posteriori* distribution of any variable X (or Y, Z, …) in the factor graph is the product $\overrightarrow{\mu}_X(x)\overleftarrow{\mu}_X(x)$ of the corresponding two messages, up to a scale factor. The parameters of this marginal distribution are m_X and W_X given by $W_X = \overrightarrow{W}_X + \overleftarrow{W}_X$ and $W_X m_X = \overrightarrow{W}_X\overrightarrow{m}_X + \overleftarrow{W}_X\overleftarrow{m}_X$.

6) Tabulated message computation rules (in particular, Tables 2–4 of [11]) allow to compose a variety of different algorithms to compute the same sum-product messages. The variety arises from different parameterizations of the messages and from local manipulations (including splitting and grouping of nodes) of the factor graph.

B. Message Passing Through $f\big(x(t_1)|x(t_0)\big)$

Gaussian message passing through the node / factor $f\big(x(t_1)|x(t_0)\big)$ is summarized in Table I. Both the forward message (with parameters (I.1) and (I.2)) and the backward message (with parameters (I.3) and (I.4)) are easily obtained from Fig. 3, (7) and (8), and Tables 2 and 3 of [11].

---

[4]The backward recursion is required for smoothing, i.e., noncausal estimation or estimation with delay; it is absent in basic Kalman filtering as in [2]. In fact, while the backward recursion is obvious from the graphical-model perspective, its development in the traditional approach was not quite so obvious, cf. [8].

If the matrix A is diagonalizable, then the integrals in (I.2) and (I.4) can easily be expressed in closed form. Specifically, if

$$A = Q\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}Q^{-1} \qquad (11)$$

for some complex square matrix Q, then

$$\int_0^t e^{A\tau}bb^\mathsf{T}e^{A^\mathsf{T}\tau}d\tau = Q\overrightarrow{\Theta}(t)Q^\mathsf{H} \qquad (12)$$

where the square matrix $\overrightarrow{\Theta}(t)$ is given by

$$\overrightarrow{\Theta}(t)_{k,\ell} \triangleq \frac{(Q^{-1}b)_k\overline{(Q^{-1}b)_\ell}}{\lambda_k + \overline{\lambda_\ell}}\left(e^{(\lambda_k+\overline{\lambda_\ell})t} - 1\right), \qquad (13)$$

and

$$\int_0^t e^{-A\tau}bb^\mathsf{T}e^{-A^\mathsf{T}\tau}d\tau = Q\overleftarrow{\Theta}(t)Q^\mathsf{H} \qquad (14)$$

with

$$\overleftarrow{\Theta}(t)_{k,\ell} \triangleq \frac{(Q^{-1}b)_k\overline{(Q^{-1}b)_\ell}}{\lambda_k + \overline{\lambda_\ell}}\left(1 - e^{-(\lambda_k+\overline{\lambda_\ell})t}\right). \qquad (15)$$

Note that, in (13) and (15), $(Q^{-1}b)_k$ denotes the k-th component of the vector $Q^{-1}b$. The proof of (12) and (14) is given in Appendix A.

The remaining entry (I.5) in Table I is Theorem 1 of the next section.

V. INPUT SIGNAL ESTIMATION AND REGULARIZED-LEAST-SQUARES INTERPRETATION

We now turn to estimating the input signal $U(t)$ and to clarifying its meaning. To this end, we need the factor graph representation of $f\big(x(t_1)|x(t_0)\big)$ that is shown in Fig. 5, which corresponds to the decomposition of (6) into N discrete steps and where $T \triangleq t_1 - t_0$. Note that this factor graph is only an approximate representation of $f\big(x(t_1)|x(t_0)\big)$, but the representation becomes exact in the limit $N \to \infty$. The variables $\tilde{U}(t)$ in Fig. 5 are related to $U(t)$ by

$$\tilde{U}(t) = \frac{N}{T}\int_{t-T/N}^t U(\tau)\,d\tau, \qquad (16)$$

i.e., $\tilde{U}(t)$ is the average of $U(t)$ over the corresponding interval. The proof of this decomposition is given in Appendix B.

For finite N, Fig. 5 is a standard linear Gaussian factor graph in which snapshots $\tilde{U}(t)$ of $U(t)$ according to (16) appear explicitly and can therefore be estimated by standard Gaussian message passing. In the resulting expression for the estimate of $\tilde{U}(t)$, we can take the limit $N \to \infty$ and thus obtain an estimate of $U(t)$.

Theorem 1. The MAP / MMSE / LMMSE estimate of $U(t)$ from observations $\tilde{Y}_k = \tilde{y}_k$ according to the system model of Section II is

$$\hat{u}(t) = \sigma_U^2 b^\mathsf{T}\left(\overrightarrow{V}_{X(t)} + \overleftarrow{V}_{X(t)}\right)^{-1}\left(\overleftarrow{m}_{X(t)} - \overrightarrow{m}_{X(t)}\right) \qquad (17)$$

where $\overrightarrow{m}_{X(t)}$, $\overrightarrow{V}_{X(t)}$, and $\overleftarrow{m}_{X(t)}$, $\overleftarrow{V}_{X(t)}$ are the parameters of the Gaussian sum-product messages as discussed in Section IV. $\qquad \square$
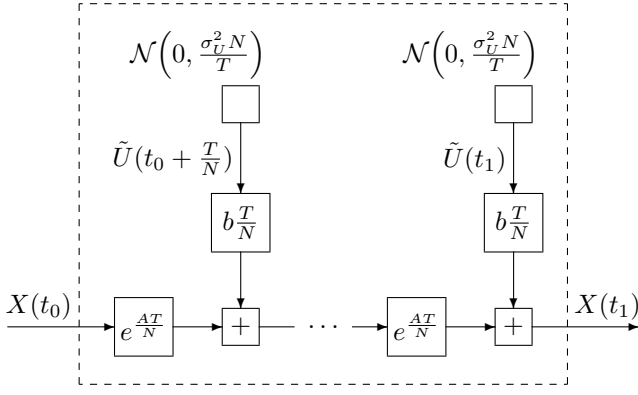
Fig. 5. Decomposition of the node/factor $f\big(x(t_1)|x(t_0)\big)$ into N discrete time steps (with $T = t_1 - t_0$). This representation is exact only in the limit $N \to \infty$.
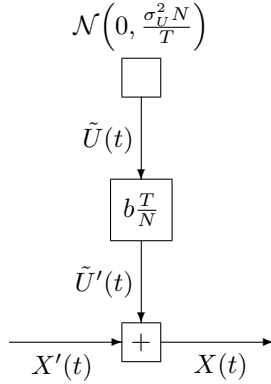

Fig. 6. Factor graph used in the proof of Theorem 1.

Proof: Consider the factor graph in Fig. 6, which shows the relevant part of Fig. 5 with suitably named variables. We determine the mean $m_{\tilde{U}(t)}$ and the variance $W_{\tilde{U}(t)}^{-1}$ of the *a posteriori* distribution of $\tilde{U}(t)$ as follows. From [11, eq. (54) and (III.5)], we have

$$W_{\tilde{U}(t)} = \overrightarrow{W}_{\tilde{U}(t)} + \overleftarrow{W}_{\tilde{U}(t)} \tag{18}$$

$$= \sigma_U^{-2}\frac{T}{N} + \left(\frac{T}{N}\right)^2 b^{\mathsf{T}}\overleftarrow{W}_{\tilde{U}'(t)}b. \tag{19}$$

From [11, eq. (55)], we then have

$$W_{\tilde{U}(t)}m_{\tilde{U}(t)} = \overrightarrow{W}_{\tilde{U}(t)}\overrightarrow{m}_{\tilde{U}(t)} + \overleftarrow{W}_{\tilde{U}(t)}\overleftarrow{m}_{\tilde{U}(t)}; \tag{20}$$

inserting $\overrightarrow{m}_{\tilde{U}(t)} = 0$ and using [11, eq. (III.6)] yields

$$W_{\tilde{U}(t)}m_{\tilde{U}(t)} = \frac{T}{N}b^{\mathsf{T}}\overleftarrow{W}_{\tilde{U}'(t)}\overleftarrow{m}_{\tilde{U}'(t)}. \tag{21}$$

Using (19) and (21), we obtain

$$m_{\tilde{U}(t)} = (W_{\tilde{U}(t)})^{-1}\left(W_{\tilde{U}(t)}m_{\tilde{U}(t)}\right) \tag{22}$$

$$= \left(\sigma_U^{-2} + \frac{T}{N}b^{\mathsf{T}}\overleftarrow{W}_{\tilde{U}'(t)}b\right)^{-1} b^{\mathsf{T}}\overleftarrow{W}_{\tilde{U}'(t)}\overleftarrow{m}_{\tilde{U}'(t)} \tag{23}$$

$$\approx \sigma_U^2 b^{\mathsf{T}}\overleftarrow{W}_{\tilde{U}'(t)}\overleftarrow{m}_{\tilde{U}'(t)} \tag{24}$$

and the approximation (24) becomes exact in the limit $N \to \infty$.

Using [11, eq. (II.10)], we have

$$\overleftarrow{m}_{\tilde{U}'(t)} = \overleftarrow{m}_{X(t)} - \overrightarrow{m}_{X'(t)} \tag{25}$$

$$\approx \overleftarrow{m}_{X(t)} - \overrightarrow{m}_{X(t)}, \tag{26}$$

and using [11, eq. (II.8)], we have

$$\overleftarrow{W}_{\tilde{U}'(t)} = \left(\overleftarrow{V}_{\tilde{U}'(t)}\right)^{-1} \tag{27}$$

$$= \left(\overrightarrow{V}_{X'(t)} + \overleftarrow{V}_{X(t)}\right)^{-1} \tag{28}$$

$$\approx \left(\overrightarrow{V}_{X(t)} + \overleftarrow{V}_{X(t)}\right)^{-1}. \tag{29}$$

Again, the approximations (26) and (29) both become exact in the limit $N \to \infty$. Inserting (26) and (29) into (24) yields

$$\lim_{N\to\infty} m_{\tilde{U}(t)} = \sigma_U^2 b^{\mathsf{T}}\left(\overrightarrow{V}_{X(t)} + \overleftarrow{V}_{X(t)}\right)^{-1}\left(\overleftarrow{m}_{X(t)} - \overrightarrow{m}_{X(t)}\right). \tag{30}$$

The mean of the *a posteriori* probability of $\tilde{U}(t)$ is thus well defined even for $N \to \infty$ and given by (30), and the theorem follows. $\square$

While we have thus established that the mean (30) of the *a posteriori* distribution of $\tilde{U}(t)$ is well-defined for $N \to \infty$, it should be pointed out that the variance of this distribution is infinite: taking the limit $N \to \infty$ of (19) yields $W_{\tilde{U}(t)} = 0$. However, this seemingly problematic result does not imply that the estimate (17) is useless; it simply reflects the obvious fact that white noise cannot be fully estimated from discrete noisy samples.

The nature of the estimate (17) is elucidated by the following theorem, which reformulates the estimation problem of this paper as an equivalent regularized least-squares problem. For the sake of clarity, we here restrict ourselves to scalar observations Y_k where $\nu = 1$, $c \triangleq C$ is a row vector, and $\sigma_Z^2 \triangleq V_Z$ is a scalar. (The general case is given in [25].)

Theorem 2. Assume that the factor graph in Fig. 2 consists of K sections between t_0 and t_K (with observations starting at t_1) and assume that the observations Y_k are scalars. Then the estimated pair $\big(\hat{u}(t), \hat{x}(t)\big)$ with $\hat{u}(t)$ as in (17) minimizes

$$\frac{1}{\sigma_U^2}\int_{t_0}^{t_K}\hat{u}(t)^2\,dt + \frac{1}{\sigma_Z^2}\sum_{k=1}^{K}\big(\tilde{y}_k - c\hat{x}(t_k)\big)^2 \tag{31}$$

subject to the constraints of the system model. $\square$

Proof: Recall the factor graph representation of a least squares problem as in Fig. 7, where the large box on top expresses the given constraints. Clearly, maximizing the function represented by Fig. 7 amounts to computing

$$\operatorname*{argmax}_{z_1,\ldots,z_n}\prod_{k=1}^{n}e^{-z_k^2/(2\sigma_k^2)} = \operatorname*{argmin}_{z_1,\ldots,z_n}\sum_{k=1}^{n}z_k^2/\sigma_k^2 \tag{32}$$

subject to the constraints. The right-hand side of (32) will be called "cost function." Recall that sum-product message passing in cycle-free linear Gaussian factor graphs maximizes the left-hand side of (32) (subject to the constraints) and thus minimizes the cost function [11].

Now plugging Fig. 5 into the factor graph in Fig. 2 results in a factor graph as in Fig. 7 with cost function
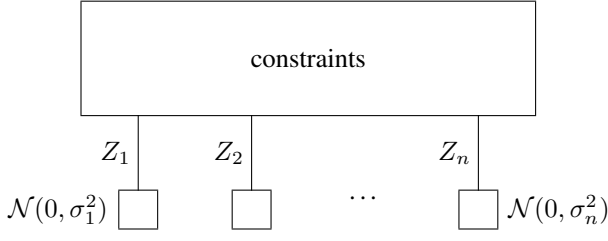
Fig. 7. Factor graphs of a least squares problem used in the proof of Theorem 2.
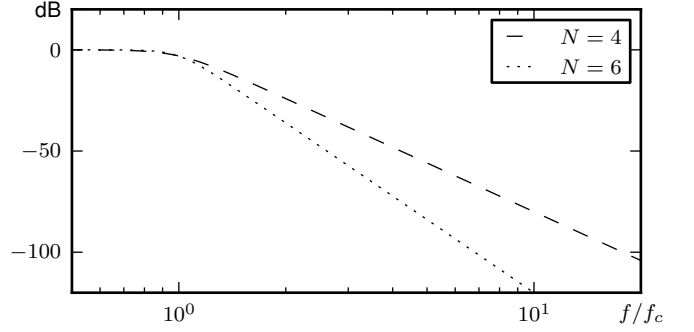

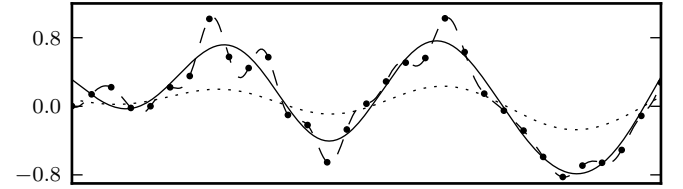Fig. 8. Frequency response (magnitude) of the filters used in Section VI.

Fig. 9. Estimation of output signal $Y(t)$ from noisy samples $\tilde{y}_k$ (fat dots) at SNR = 10 dB. Solid line: estimate of $Y(t)$ at correct SNR. Dashed line: estimation with assumed SNR 100 dB; dotted line: estimation with assumed SNR -10 dB.

$$\sum_{k=1}^{K}\left(z_k^2/\sigma_Z^2 + \sum_{\ell=1}^{N}\tilde{u}\Big(t_{k-1}+\ell\frac{T_k}{N}\Big)^2\frac{T_k}{\sigma_U^2 N}\right)$$
$$= \sum_{k=1}^{K}\left(z_k^2/\sigma_Z^2 + \frac{1}{\sigma_U^2}\int_{t_{k-1}}^{t_k}u(t)^2\,dt\right), \quad (33)$$

which is (31). □

According to Theorem 2, minimizing (31) is mathematically equivalent to the statistical estimation problems of this paper; in particular, modeling $U(t)$ as white Gaussian noise amounts to regularizing the second term in (31) by penalizing power in $\hat{u}(t)$.

The functional (31) is amenable to an informal frequency-domain analysis that considers the relative power in the different frequences of the input signal $\hat{u}(t)$. In particular, the estimate $\hat{u}(t)$ fits the corresponding output signal $\hat{y}(t) = c\hat{x}(t)$ to the observations $\tilde{y}_k$ preferably by those frequencies that appear with little damping in the output signal. Since the transfer function from $U(t)$ to $Y(t) = cX(t)$ of the system (1) is necessarily a (non-ideal) low-pass filter, the estimate $\hat{u}(t)$ will contain little energy in very high frequencies. In this way, the spectrum of $\hat{u}(t)$ is shaped by the transfer function of the linear system.

We also note[5] that the problem of minimizing (31) may be viewed as an offline control problem where an input signal $u(t)$ is to be determined such that the resulting sampled output signal $y_1, y_2, \ldots$ follows a desired trajectory $\tilde{y}_1, \tilde{y}_2, \ldots$. However, exploring this connection to control theory is beyond the scope of this paper.

VI. NUMERICAL EXAMPLES

We illustrate the estimators of this paper by some simple numerical examples. In all these examples, the output signal $Y(t)$ is scalar, we use regular sampling at rate f_s, i.e., $Y_k = Y(k/f_s)$, and the linear system in Fig. 1 is a Butterworth lowpass filter of order 4 or 6 with cut-off frequency (-3 dB frequency) f_c [27]. The amplitude response (i.e., the magnitude of the frequency response) of these filters is plotted in Fig. 8.

In these examples, we use the signal-to-noise ratio (SNR) as discussed in Appendix C. Using (57), the SNR of the discrete-time observations turns out to be

$$\text{SNR} \approx \frac{\sigma_U^2}{\sigma_Z^2}f_c\cdot 2.052 \quad (34)$$

---
[5]This was pointed out to the authors by Andrew Singer of the University of Illinois at Urbana-Champain.

for the 4th-order filter and

$$\text{SNR} \approx \frac{\sigma_U^2}{\sigma_Z^2}f_c\cdot 2.023 \quad (35)$$

for the 6th-order filter. We will measure the SNR in dB (i.e., $10\cdot\log_{10}(\text{SNR})$).

In some of these plots, the estimator deliberately assumes an incorrect SNR, i.e., an incorrect ratio σ_U^2/σ_Z^2, in order to illustrate the effect of this ratio on (31).

Estimation of the output signal $Y(t)$ is illustrated in Fig. 9. In this example, the linear system is a Butterworth filter of order 6. The noisy samples $\tilde{y}_k$ are created with $f_s = 10f_c$ at an SNR of 10 dB. The corresponding estimate of $Y(t)$ is shown as solid line in Fig. 9.

Also shown in Fig. 9 is the effect of estimating with an incorrect SNR, i.e., of playing with the ratio σ_U^2/σ_Z^2 as mentioned above. Estimating with an assumed SNR that is too high results in overfitting; estimating with an assumed SNR that is too low reduces the amplitude of the estimated signal.

Fig. 10 shows the effect of f_s/f_c on the normalized estimation error

$$\text{SNR}_{\text{out}}^{-1} \triangleq \frac{\text{E}\left[(\hat{Y}_k - Y_k)^2\right]}{\text{E}[Y_k^2]} \quad (36)$$

for a Butterworth filter of order 4. For high SNR, we clearly see a critical "Nyquist region" where severe undersampling sets in. For large f_s/f_c, the estimate improves by about 2.62 dB with every factor of 2 in f_s/f_c, which is less than what would be expected (viz., 3 dB) for strictly bandlimited signals [28], [29].

Estimation of the input signal $U(t)$ is illustrated in Fig. 11, for exactly the same setting (with the same discrete-time
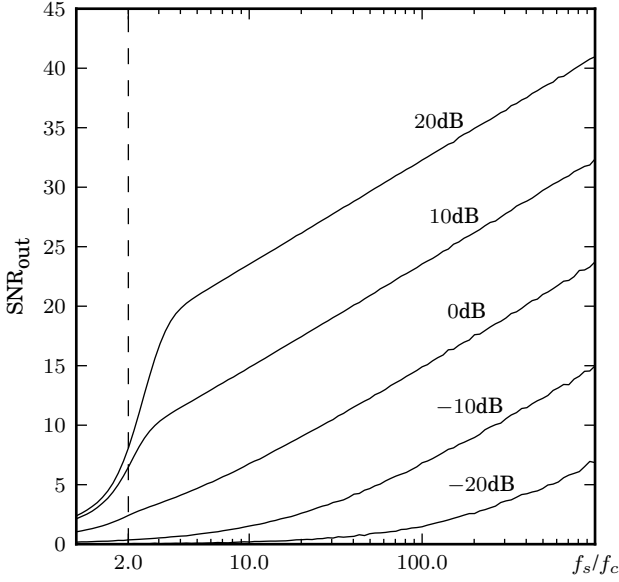
Fig. 10. Empirical estimation error (36) vs. normalized sampling frequency f_s/f_c, parameterized by the SNR (54), for a Butterworth filter of order 4.
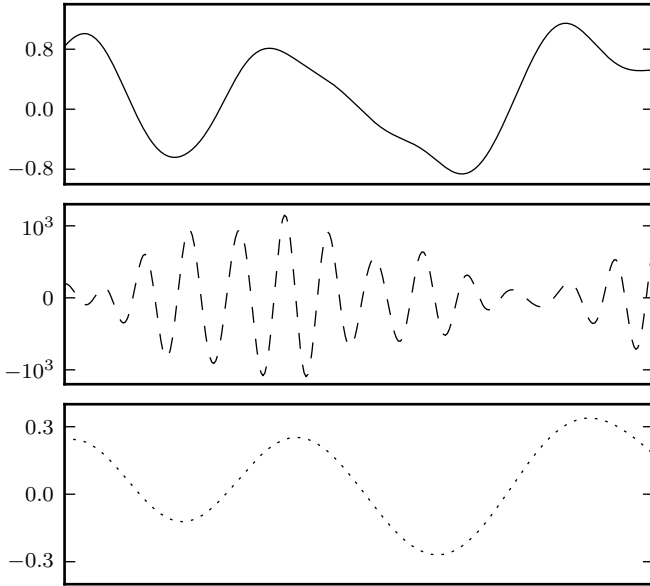

Fig. 11. Input signal estimation for the same cases (and the same time scale) as in Fig. 9. The solid line (top) is the correct MMSE/LMMSE estimate of $U(t)$.

observations $\tilde{y}_k$) as in Fig. 9. The power and the spectral content for the three different plots in Fig. 11 illustrate the effect of the ratio σ_U^2/σ_Z^2 on (31).

VII. EXTENSIONS

We briefly mention a number of extensions and modifications of the system model that are required in some of the motivating applications and are easily incorporated in the estimation algorithms.

A. Additional Spectral Shaping

The estimate (17) of the input signal $U(t)$ is marked by an implicit spectral shaping (cf. the discussion after Theorem 2).

It may sometimes be desirable, however, to control the spectrum of the estimate more explicitly. This can be achieved by assuming that the input signal $U(t)$ is not white Gaussian noise, but white Gaussian noise passed through a suitable (finite-dimensional) linear prefilter. The estimation of $U(t)$ is easily adapted to this case by including the prefilter in the system model.

In contrast to unfiltered-input estimation as in Section V, estimation of a filtered input signal by means of Kalman filtering / smoothing is standard.

B. Time-Varying and Affine Systems

In some applications, the dynamics of the system / filter in Fig. 1 may change at discrete instants in time (but it is always known). This situation occurs, e.g., when the analog system / filter is subject to digital control. An example of such a case is given in [18].

We thus generalize the system model (5) and (2) to

$$dX(t) = \big(A_k X(t) + b_k U(t) + h_k\big)\, dt \qquad (37)$$

and

$$Y_k = C_k X(t_k), \qquad (38)$$

which holds for $t_k \le t < t_{k+1}$, where A_k and C_k are known matrices, and where b_k and h_k are known column vectors.

If $h_k = 0$, both the factor graph representations and the message computation rules remain unchanged except for the addition of subscripts to the involved matrices and vectors. The case $h_k \ne 0$ is included below.

C. Multiple Inputs and Internal Noise

We are also interested in the case where the system / filter in Fig. 1 has internal noise sources. (Again, a main motivation are analog-to-digital converters, where the noise in the analog part cannot be neglected.) Such internal noise can be handled mathematically by extending the input signal $U(t)$ to a vector $U(t) = \big(U_1(t), U_2(t), \dots\big)^{\mathsf{T}}$, where the first component, $U_1(t)$, is the actual input signal while the remaining components model the internal noise. For $t < t'$, the integral

$$\int_t^{t'} U(\tau)\, d\tau \qquad (39)$$

is a zero-mean Gaussian random vector with diagonal covariance matrix $\sigma_U^2 I(t' - t)$. The corresponding generalization of (5) is

$$dX(t) = \big(AX(t) + BU(t) + h\big)\, dt, \qquad (40)$$

where B is a matrix of suitable dimensions and where we have included a constant offset h (a column vector) as in (37). Note that power differences and correlations among the input signals can be expressed by a suitable matrix B.

The corresponding generalization of Table I is shown in Table II. The proofs are straightforward modifications of the proofs of Table I and are omitted.

TABLE II
GENERALIZATION OF TABLE I TO (40).

$$X(t_0) \quad \boxed{} \quad X(t_1)$$

$$f\big(x(t_1)|x(t_0)\big)$$

$$\overrightarrow{m}_{X(t_1)} = e^{A(t_1-t_0)}\overrightarrow{m}_{X(t_0)} + A^{-1}\big(e^{A(t_1-t_0)} - I\big)h \quad \text{(II.1)}$$

$$\overrightarrow{V}_{X(t_1)} = e^{A(t_1-t_0)}\overrightarrow{V}_{X(t_0)}e^{A^\mathsf{T}(t_1-t_0)}$$
$$+ \sigma_U^2 \underbrace{\int_0^{t_1-t_0} e^{A\tau}BB^\mathsf{T}e^{A^\mathsf{T}\tau}d\tau}_{Q\overrightarrow{\Theta}(t_1-t_0)Q^\mathsf{H} \quad \text{see (41)}} \quad \text{(II.2)}$$

$$\overleftarrow{m}_{X(t_0)} = e^{-A(t_1-t_0)}$$
$$\Big(\overleftarrow{m}_{X(t_1)} - A^{-1}\big(e^{A(t_1-t_0)} - I\big)h\Big) \quad \text{(II.3)}$$

$$\overleftarrow{V}_{X(t_0)} = e^{-A(t_1-t_0)}\overleftarrow{V}_{X(t_1)}e^{-A^\mathsf{T}(t_1-t_0)}$$
$$+ \sigma_U^2 \underbrace{\int_0^{t_1-t_0} e^{-A\tau}BB^\mathsf{T}e^{-A^\mathsf{T}\tau}d\tau}_{Q\overleftarrow{\Theta}(t_1-t_0)Q^\mathsf{H} \quad \text{see (42)}} \quad \text{(II.4)}$$

$$\hat{u}(t) = \sigma_U^2 B^\mathsf{T}\big(\overrightarrow{V}_{X(t)} + \overleftarrow{V}_{X(t)}\big)^{-1}\big(\overleftarrow{m}_{X(t)} - \overrightarrow{m}_{X(t)}\big) \quad \text{(II.5)}$$

If the matrix A is diagonalizable as in (11), then the integrals in (II.2) and (II.4) can be written as stated in the table where the square matrices $\overrightarrow{\Theta}(t)$ and $\overleftarrow{\Theta}(t)$ are given by

$$\overrightarrow{\Theta}(t)_{k,\ell} \triangleq \frac{\psi_{k,\ell}}{\lambda_k + \overline{\lambda_\ell}}\Big(e^{(\lambda_k+\overline{\lambda_\ell})t} - 1\Big) \quad \text{(41)}$$

and by

$$\overleftarrow{\Theta}(t)_{k,\ell} \triangleq \frac{\psi_{k,\ell}}{\lambda_k + \overline{\lambda_\ell}}\Big(1 - e^{-(\lambda_k+\overline{\lambda_\ell})t}\Big), \quad \text{(42)}$$

respectively, and where $\psi_{k,\ell}$ is the entry in row k and column ℓ of the matrix

$$\Psi \triangleq Q^{-1}B\,(Q^{-1}B)^\mathsf{H}. \quad \text{(43)}$$

D. Nonlinearities

Mild nonlinearities in the system / filter in Fig. 1 can often be handled by extended Kalman filtering [7], [8], i.e., by iterative estimation using a linearized model based on a tentative estimate of the state trajectory $X(t)$.

VIII. CONCLUSIONS

We have pointed out that exact models of continuous-time linear systems driven by white Gaussian noise can be used in discrete-time factor graphs. The associated continuous-time signals then become computational objects that can be handled with arbitrary temporal resolution by discrete-time Gaussian message passing.

Motivated by applications such as dynamical sensors and analog-to-digital converters, we have been particularly interested in estimating the input signal, which does not seem to have been addressed in the prior Kalman filtering literature.
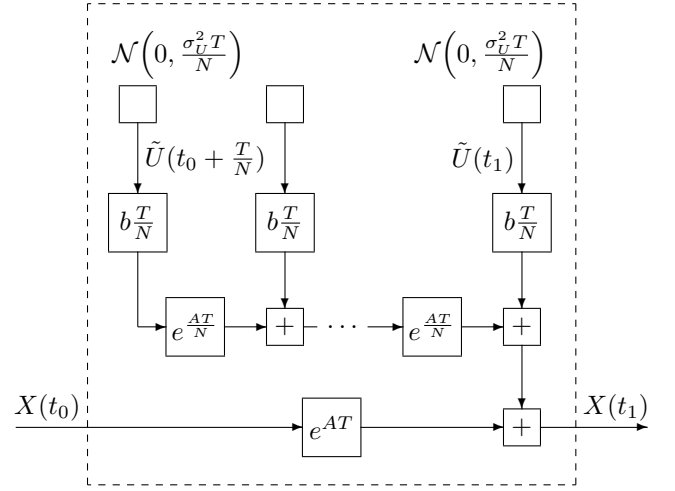

Fig. 12. Decomposition of the node / factor $f\big(x(t_1)|x(t_0)\big)$ into N discrete time steps according to (53).

APPENDIX A
PROOF OF (12) AND (14)

Let

$$\Lambda \triangleq \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}. \quad \text{(44)}$$

From (11), we have

$$e^{A\tau} = Qe^{\Lambda\tau}Q^{-1} \quad \text{(45)}$$

and

$$e^{A^\mathsf{T}\tau} = (e^{A\tau})^\mathsf{T} = (e^{A\tau})^\mathsf{H} = (Q^{-1})^\mathsf{H}e^{\overline{\Lambda}\tau}Q^\mathsf{H}, \quad \text{(46)}$$

and thus

$$\int_0^t e^{A\tau}bb^\mathsf{T}e^{A^\mathsf{T}\tau}d\tau = Q\left(\int_0^t e^{\Lambda\tau}\Psi e^{\overline{\Lambda}\tau}\,d\tau\right)Q^\mathsf{H} \quad \text{(47)}$$

with

$$\Psi \triangleq Q^{-1}b(Q^{-1}b)^\mathsf{H}. \quad \text{(48)}$$

The element in row k and column ℓ of the matrix under the integral is

$$\Big(e^{\Lambda\tau}\Psi e^{\overline{\Lambda}\tau}\Big)_{k,\ell} = \psi_{k,\ell}\,e^{(\lambda_k+\overline{\lambda_\ell})\tau}, \quad \text{(49)}$$

where $\psi_{k,\ell}$ refers to the elements of the matrix Ψ, and elementwise integration yields

$$\left(\int_0^t e^{\Lambda\tau}\Psi e^{\overline{\Lambda}\tau}\,d\tau\right)_{k,\ell} = \frac{\psi_{k,\ell}}{\lambda_k + \overline{\lambda_\ell}}\Big(e^{(\lambda_k+\overline{\lambda_\ell})t} - 1\Big), \quad \text{(50)}$$

which proves (12). The proof of (14) follows from noting that changing $e^{A\tau}$ into $e^{-A\tau}$ amounts to a sign change of Λ.
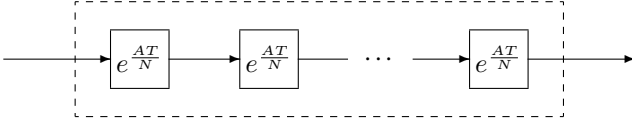
Fig. 13. Decomposition of e^{AT} into N sections.

APPENDIX B
PROOF OF THE DISCRETE-TIME DECOMPOSITION IN FIG. 5

We split the integral (6) into N parts, each of width T/N with $T \triangleq t_1 - t_0$:

$$X(t_1) = e^{AT} X(t_0)$$
$$+ \sum_{k=1}^{N} \int_{(k-1)T/N}^{kT/N} e^{A(T-\tau)} b U(t_0 + \tau) \, d\tau \quad (51)$$

$$\approx e^{AT} X(t_0)$$
$$+ \sum_{k=1}^{N} e^{A(T-kT/N)} b \int_{(k-1)T/N}^{kT/N} U(t_0 + \tau) \, d\tau \quad (52)$$

$$= e^{AT} X(t_0)$$
$$+ \sum_{k=1}^{N} e^{A(T-kT/N)} b \frac{T}{N} \tilde{U}(t_0 + kT/N), \quad (53)$$

where the approximation (52) becomes exact in the limit $N \to \infty$ and where $\tilde{U}(t)$ is defined as in (16). The factor graph of (53) is shown in Fig. 12.

The term e^{AT} can also be decomposed into N discrete steps as shown in Fig. 13. Plugging Fig. 13 into Fig. 12 yields a factor graph that is easily seen to be equivalent to Fig. 5.

APPENDIX C
ON SNR

For the system model of Section II, we may wish to relate the input noise power σ_U^2 to the signal-to-noise ratio (SNR) of the discrete-time observations. For the sake of clarity, we restrict ourselves to scalar observations Y_k, i.e., $\nu = 1$, $c \triangleq C$ is a row vector, and $\sigma_Z^2 \triangleq V_Z$ is a scalar. In addition, we assume that the continuous-time linear system is time-invariant and stable and any initial conditions can be neglected. In this case, we define

$$\text{SNR} \triangleq \frac{\mathrm{E}[Y_k^2]}{\sigma_Z^2} \quad (54)$$

which (under the stated assumptions) is independent of k. We then have

$$\mathrm{E}[Y_k^2] = c \overrightarrow{V}_{X(\infty)} c^{\mathsf{T}} \quad (55)$$

with

$$\overrightarrow{V}_{X(\infty)} \triangleq \sigma_U^2 \lim_{t \to \infty} \int_0^t e^{A\tau} b b^{\mathsf{T}} e^{A^{\mathsf{T}}\tau} d\tau \quad (56)$$

from (I.2); if, in addition, the system is diagonalizable as in (11), then

$$\mathrm{E}[Y_k^2] = \sigma_U^2 c Q \overrightarrow{\Theta}(\infty) Q^{\mathsf{H}\mathsf{T}} c^{\mathsf{T}} \quad (57)$$

where $\overrightarrow{\Theta}(\infty)$ is a square matrix with entries

$$\overrightarrow{\Theta}(\infty)_{k,\ell} = -\frac{(Q^{-1}b)_k \overline{(Q^{-1}b)_\ell}}{\lambda_k + \overline{\lambda_\ell}} \quad (58)$$

REFERENCES

[1] L. Bolliger, H.-A. Loeliger, and C. Vogel, "Simulation, MMSE estimation, and interpolation of sampled continuous-time signals using factor graphs," 2010 Information Theory & Applications Workshop, UCSD, La Jolla, CA, USA, Jan. 31 – Feb. 5, 2010.

[2] R. E. Kalman, "A new approach to linear filtering and prediction problems," Trans. ASME, Series D, J. of Basic Eng., vol. 82, 1960, pp. 35–45.

[3] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," Trans. ASME, Series D, J. of Basic Eng., vol. 83, 1961, pp. 95–107.

[4] B. D. O. Anderson and J. B. Moore, Optimal Filtering. Prentice Hall, NJ, 1979.

[5] R. J. Meinhold and N. D. Singpurwalla, "Understanding the Kalman filter," American Statistician, vol. 37, no. 2, pp. 123–127, May 1983.

[6] S. Haykin, Adaptive Filter Theory, 3rd ed. Prentice Hall, NJ, 1996.

[7] M. S. Grewal and A. P. Andrews, Kalman Filtering: Theory and Practice Using MATLAB. 2nd ed., Wiley 2001.

[8] T. Kailath, A. H. Sayed, and B. Hassibi, Linear Estimation. Prentice Hall, NJ, 2000.

[9] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," IEEE Trans. Inform. Theory, vol. 47, pp. 498–519, Feb. 2001.

[10] H.-A. Loeliger, "An introduction to factor graphs," IEEE Signal Proc. Mag., Jan. 2004, pp. 28–41.

[11] H.-A. Loeliger, J. Dauwels, Junli Hu, S. Korl, Li Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," Proceedings of the IEEE, vol. 95, no. 6, pp. 1295–1322, June 2007.

[12] H. Wymeersch, Iterative Receiver Design. Cambridge University Press, 2007.

[13] M. I. Jordan, "Graphical models," Statistical Science, vol. 19, no. 1, pp. 140–155, 2004.

[14] Ch. M. Bishop, Pattern Recognition and Machine Learning. New York: Springer Science+Business Media, 2006.

[15] M. J. Wainwright and M. I. Jordan, "Graphical Models, Exponential Families, and Variational Inference," Foundations and Trends in Machine Learning, vol. 1, no. 1–2, pp.1–305, 2008.

[16] D. Koller and N. Friedman, Probabilistic Graphical Models. Cambridge, MA, MIT Press, 2009.

[17] L. Bolliger and H.-A. Loeliger, "Sampling jitter correction using factor graphs," Proc. 2011 Europ. Signal Proc. Conf. (EUSIPCO), Barcelona, Spain, Aug. 29 – Sept. 2, 2011.

[18] H.-A. Loeliger, L. Bolliger, G. Wilckens, and J. Biveroni, "Analog-to-digital conversion using unstable filters," 2011 Information Theory & Applications Workshop, UCSD, La Jolla, CA, USA, Feb. 6–11, 2011.

[19] U. Nodelman, C. R. Shelton, and D. Koller, "Continuous-time Bayesian networks, Proc. 18th Conf. on Uncertainty in Artificial Intell., pp. 378–387, 2002.

[20] T. El-Hay, I. Cohn, N. Friedman, and R. Kupferman, "Continuous-time belief propagation," Proc. 27th Int. Conf. on Machine Learning, Haifa, Israel, June 21–24, 2010.

[21] I. Cohn, T. El-Hay, N. Friedman, and R. Kupferman, "Mean field variational approximation for continuous-time Bayesian networks," J. Machine Learning Research, vol. 11, pp. 2745–2783, 2010.

[22] P. O. Vontobel and H.-A. Loeliger, "Factor graphs and dynamical electrical networks," Proc. 2003 IEEE Information Theory Workshop, Paris, France, March 31 – April 4, 2003, pp. 218–221.

[23] H. Garnier and L. Wang, Identification of Continuous-Time Models from Sampled Data. Springer Verlag 2008.

[24] T. Söderström, Discrete-Time Stochastic Systems, 2nd ed. London: Springer Verlag, 2002.

[25] L. Bolliger, Digital Estimation of Continuous-Time Signals Using Factor Graphs. PhD thesis no. 20123 at ETH Zurich, 2012.

[26] G. D. Forney, Jr., "Codes on graphs: normal realizations," IEEE Trans. Inform. Theory, vol. 47, no. 2, pp. 520–548, 2001.

[27] A. Oppenheim and A. Willsky, Signals and Systems, 2nd ed. Prentice Hall, 1996.

[28] W. R. Bennet, "Spectra of quantized signals," Bell Syst. Techn. J., vol. 27, pp. 446–472, July 1948.

[29] N. T. Thao and M. Vetterli, "Lower bound on the mean-squared error in oversampled quantization of periodic signals using vector quantization analysis," IEEE Trans. Inform. Theory, vol. 42, no. 2, pp. 469–479, March 1996.